Techniques for Large-Scale Automatic Detection of Web Site Defacements

Eric Medvet

Abstract

Web site defacement, the process of introducing unauthorized modifications to a web site, is a very common form of attack. This thesis describes the design and experimental evaluation of a framework that may constitute the basis for a defacement detection service capable of monitoring thousands of remote web sites systematically and automatically. With this framework an organization may join the service by simply providing the URL of the resource to be monitored along with the contact point of an administrator. The monitored organization may thus take advantage of the service with just a few mouse clicks, without installing any software locally nor changing its own daily operational processes.

The main proposed approach is based on anomaly detection and allows monitoring the integrity of many remote web resources automatically while remaining fully decoupled from them, in particular, without requiring any prior knowledge about those resources. During a preliminary learning phase a profile of the monitored resource is built automatically. Then, while monitoring, the remote resource is retrieved periodically and an alert is generated whenever something "unusual" shows up.

The thesis discusses about the effectiveness of the approach in terms of accuracy of detection—i.e., missed detections and false alarms. The thesis also considers the problem of misclassified readings in the learning set. The effectiveness of anomaly detection approach, and hence of the proposed framework, bases on the assumption that the profile is computed starting from a learning set which is not corrupted by attacks; this assumption is often taken for granted. The influence of leaning set corruption on our framework effectiveness is assessed and a procedure aimed at discovering when a given unknown learning set is corrupted by positive readings is proposed and evaluated experimentally.

An approach to automatic defacement detection based on Genetic Programming (GP), an automatic method for creating computer programs by means of artificial evolution, is proposed and evaluated experimentally. Moreover, a set of techniques that have been used in literature for designing several host-based or network-based Intrusion Detection Systems are considered and evaluated experimentally, in comparison with the proposed approach.

Finally, the thesis presents the findings of a large-scale study on reaction time to web site defacement. There exist several statistics that indicate the number of incidents of this sort but there is a crucial piece of information still lacking: the typical duration of a defacement. A two months monitoring activity has been performed over more than 62000 defacements in order to figure out whether and when a reaction to the defacement is taken. It is shown that such time tends to be unacceptably long—in the order of several days—and with a long-tailed distribution.